Genomics 93 (2009) 169–178

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Characterization of the genome expression trends in the heading-stage panicle of six rice lineages

Zhi-Yu Peng [a,b,1], Huiyong Zhang [c,d,1], Tingting Liu [a,c], Katherine M. Dzikiewicz [e], Songgang Li [b], Xiangfeng Wang [d], Guocheng Hu [f], Zhengge Zhu [g], Xinghua Wei [f], Qi-Hui Zhu [b,h], Zongxiu Sun [f], Song Ge [h], Ligeng Ma [c], Lei Li [e,*], Xing-Wang Deng [a,c,d,*]

[a] Peking-Yale Joint Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing, 100871 People's Republic of China
[b] Center for Bioinformatics, National Laboratory of Protein Engineering and Plant Genetic Engineering, College of Life Sciences, Peking University, Beijing 100871, People's Republic of China
[c] National Institute of Biological Sciences, 7 Science Park Road, Zhongguancun Life Science Park, Beijing, 102206 People's Republic of China
[d] Department of Molecular, Cellular, and Developmental Biology, Yale University New Haven, CT 06520, USA
[e] Department of Biology, University of Virginia, Charlottesville VA 22904, USA
[f] State Key Laboratory of Rice Biology, China National Rice Research Institute, 359 Tiyuchang Road, 310006 Hangzhou, Zhejiang Province, People's Republic of China
[g] College of Life Sciences, Hebei Normal University, Shijiazhuang 050016, China
[h] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, People's Republic of China

ARTICLE INFO

Article history:
Received 14 August 2008
Accepted 7 October 2008
Available online 25 November 2008

Keywords:
Gene expression
Comparative genomics
Microarray
Rice
Domestication

ABSTRACT

To study how changes in gene regulation shape phenotypic variations in rice, we performed a comparative analysis of genome expression in the heading-stage panicle from six lineages of cultivated and wild rice, including *Oryza* sativa subsp. indica, japonica and javanica, O. nivara , O. rufipogon and O. glaberrima. While nearly three-quarters of the genes are expressed at a constant level in all six lineages, a large portion of the genome, ranging from 1767 to 4489 genes, exhibited differential expression between Asian domesticated and wild rice with repression or down-regulation of genome expression in Asian cultivated rice as the dominant trend. Importantly, we found this repression was achieved to a large extent by the differential expression of a single member of paralogous gene families. Functional analysis of the differentially expressed genes revealed that genes related to catabolism are repressed while genes related to anabolism up-regulated. Finally, we observed that distinct evolutionary forces may have acted on gene expression and the coding sequences in the examined rice lineages.

Published by Elsevier Inc.

Introduction

It was recognized decades ago by genetists that many phenotypic changes between closely related species are due to differences in gene expression regulation rather than in the coding sequences [1]. Still, relatively little is known how natural selection acts on gene regulation on the whole genome scale. One approach to study this problem has been to examine how regulatory sequences evolve [2,3]. However, regulatory sequences on a genome scale are difficult to identify [4,5] and hence their sequence variations difficult to characterize. The very nature of complex traits suggests that variation at multiple loci is required to generate a phenotype, which poses another formidable problem for associating sequence variation with complex phenotypes. Furthermore, the relation between regulatory sequence variation and gene regulation is unclear for most genes that have not been subject to detailed molecular study. In at least some cases, gene regulation is found to be conserved even when the putative regulatory sequences have changed [6–8].

In recent years, microarray-based analysis, which permits simultaneous measurement of the expression of multiple genes, provided an alternative approach to study the genetic basis of phenotypic variation [9]. Multiplex and accurate measurement of transcript levels allows gene expression patterns to be treated with quantitative genetic principles that facilitate the study of how natural selection shapes gene expression variation between related species on a genome-wide scale [9]. This type of study has been carried out in several model species including primates [10–13], flies [14,15], mice [16], and worms [17]. Focus of these studies was primarily on characterizing the general trends in the evolution of gene expression. However, our knowledge is still limited in this field, especially with regard to gene expression evolution in higher plants.

Domesticated rice (*Oryza sativa*) is one of the most important staple food crops feeding half the world's population. This together

Abbreviations: BAC, bacterial artificial chromosome; DEG, differentially expressed gene.

* Corresponding authors. L. Li is to be contacted at Department of Biology, University of Virginia, Charlottesville VA 22904, USA. X.-W. Deng, Department of Molecular, Cellular, and Developmental Biology, Yale University New Haven, CT 06520, USA. Fax: +86 203 432 3854.
E-mail addresses: ll4jn@virginia.edu (L. Li), xingwang.deng@yale.edu (X.-W. Deng).
[1] These authors contributed equally to this work.

0888-7543/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.ygeno.2008.10.005

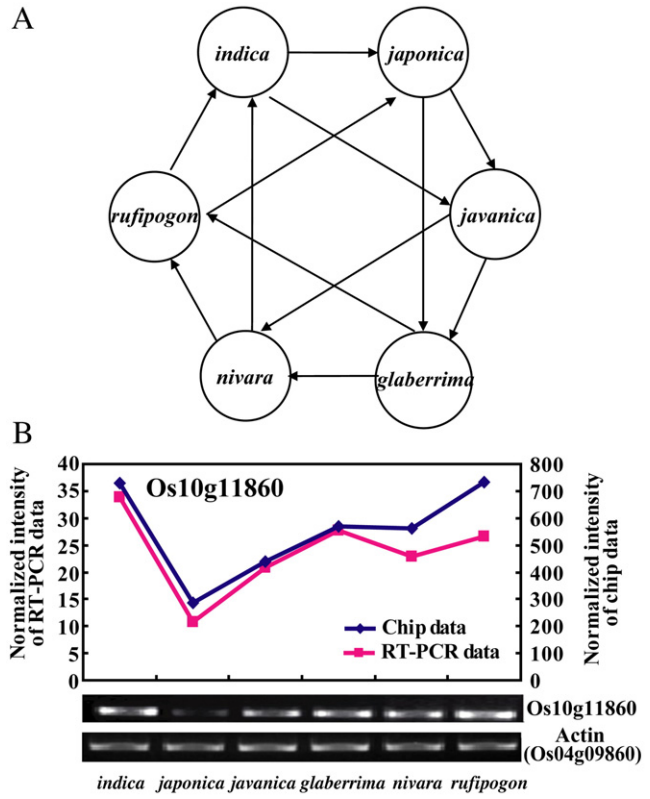

Fig. 1. Microarray analysis of gene expression pattern in six rice lineages. (A) Loop-design of the microarray experiments. Each arrow connects a pair of samples used in the same hybridization. The arrowheads point to Cy3-labeled samples while the arrow ends represent the Cy5-labeled samples. (B) Verification of microarray analysis by semi-quantitative RT-PCR. For each gene, the normalized array intensities in *indica*, *japonica*, *javanica*, *glaberrima*, *nivara* and *rufipogon* were plotted against the RT-PCR product intensities normalized against the rice actin gene (Os04g09860).

with the availability of complete [18–20] and partial genome sequences [21] in several related rice species makes it an important system to study gene expression evolution. Furthermore, rice provides a linchpin in comparative genomics aimed at elucidating genome organization and evolution in diverse monocotyledonous species. Thus, knowledge of gene expression changes that underlie phenotypic variation will be transferable to other food and feed crops.

Domestication is a long process of selecting traits that resulted in major alterations of plant structure and reproductive physiology [22]. Panicle morphology is one of the main determinants of rice yield [23] and thus exhibits great variations between cultivated rice and wild rice [24]. Compared to wild rice, cultivated rice displays an increased synchronization of panicle formation, more secondary panicle branches [22], and more densely packed panicles that can carry larger numbers of seeds than the wild ancestors [25]. Our knowledge about the genetic differences of panicles between cultivated and wild rice is still limited. In particular, the heading stage connects the reproductive stage (from panicle initiation to heading) and the grain filling or ripening stage (from heading to maturity), and is a key process during rice plant growth [26]. Therefore, characterization of the global gene expression patterns in the heading-stage panicle of related rice lineages will have the corollary benefit of identifying the biochemical and genetic pathways that contribute to yield increase in cultivated rice.

Here we report a comparative analysis of genome expression of the panicle tissue in the heading-stage in six different rice lineages. We found that while a majority of the genes are expressed at a constant level among all six rice lineages, significant decreases in genome expression were found in cultivated rice lineages. Further analysis of the metabolic pathways associated with the differentially expressed genes (DEGs) among the examined lineages revealed a general trend that expression of genes related to catabolism and anabolism was repressed and increased in cultivated lineages, respectively. We also found that a significant source for the DEGs was distinct members of the paralogous gene families. Finally, we observed that different evolutionary forces may have acted on gene expression and the coding sequences.

Results

Assess global gene expression patterns in six rice lineages

Towards understanding gene expression variation among related rice lineages on a genome-wide scale, we sought to assess global gene expression in the heading-stage panicle using a whole genome oligonucleotide microarray designed to represent 36,926 annotated *indica* genes [27]. Using a loop-design, we examined gene expression patterns in six related rice lineages (Fig. 1A), including O. sativa (three Asian cultivars indica, japonica and javanica), O. nivara (Asian annual wild rice), O. rufipogon (Asian perennial wild rice) and O. glaberrima (African cultivated rice).

Although the array probes may not match perfectly to the targeted genes in other rice lineages due to sequence divergence, several lines of evidence indicate that this array is suitable for examining the rice lineages involved in our experiment. First, mapping the oligonucleotide probe sequences to available genome sequences showed that regions represented by the probes are highly conserved with virtually all probes exhibiting >80% identify to its target (Fig. S1). Though *javanica* was not analyzed due to the lack of sequences information, we estimated that its sequence mismatch degree between probe and targets is highly similar to *japonica* as they show extraordinarily similar phenotypes that some consider *javanica* and *japonica* to be a single species [28]. On the other hand, *nivara* and *rufipogon*, considered by some to be a single species due to their highly similar morphology [29,30], showed virtually the same degree of mismatch (Fig. S1). Second, we carried out RT-PCR analysis of the expression of 60 randomly selected genes in the six rice lineages. This experiment confirmed the general expression pattern across the lineage as measured by microarrays for 57 (95%) genes (Fig. S2). One such example (Os10g11860) is illustrated in Fig. 1B. Finally, similar results were obtained using only the perfect-match probes for the key analysis (see below). Together, these results indicate that the microarray analysis provides a reliable measurement of global gene expression in the rice lineages involved in our study.

Trends of genome expression changes in the six rice lineages

We first identified genes that were expressed at constant levels in all six rice lineages. It's notable that a majority of the genes (26,802 or 72.6%) has constant expression levels in all six rice lineages. Even using more stringent criteria, we could still identify 16,288 (44.1%) such genes. The fact that most gene expression remained constant is consistent with the previous studies from other model organisms [11,14,31]. Functional analysis of the Gene Ontology terms associated with the mentioned 16,288 genes showed significant enrichment for development, signal regulation, transcription activity and various kinds of plant defense responses (Table S1).

Table 1
Number of DEGs among examined rice lineages

	indica	*japonica*	*javanica*	*nivara*	*rufipogon*
japonica	5116				
javanica	3121	2670			
nivara	1767	4489	3008		
rufipogon	4159	4178	1884	5105	
glaberrima	2415	3848	3582	2720	4713

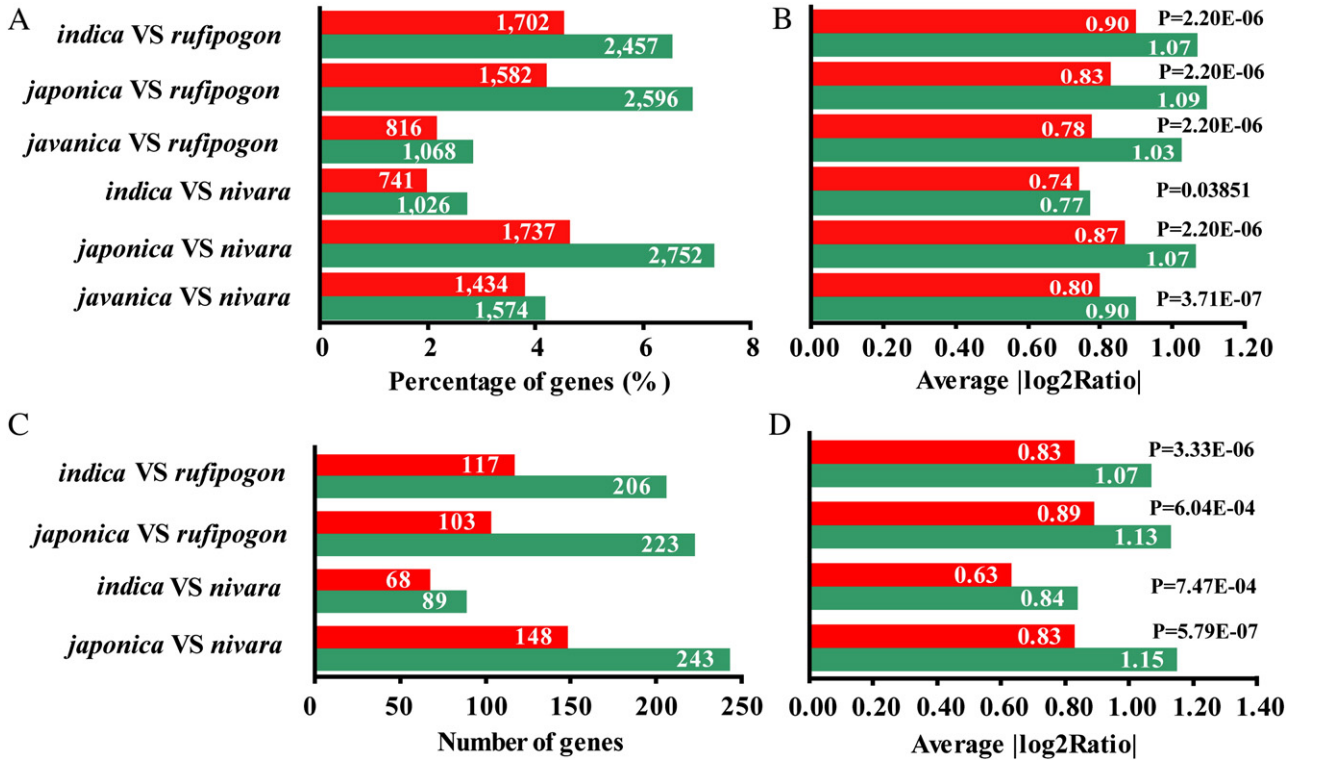

Fig. 2. Number and expression level of DEGs between Asian cultivated rice and Asian wild rice. (A) The number and percentage of genes expressed differentially between Asian cultivated rice and Asian wild rice. Red indicates up-regulated genes and green down-regulated genes. (B) The average absolute value of log2 ratio of genes expressed differentially between Asian cultivated rice and Asian wild rice. The *p*-value from Wilcox test with the null hypothesis that up-regulated genes have the larger mean than down regulated-genes was shown for each comparison. (C and D) The same analysis as shown in A and B, respectively, with only perfect-match probes among all lineages.

We then identified genes expressed differentially among the six examined rice lineages. This analysis identified a large portion of the genome, ranging from 1767 to 5116 genes, exhibited differential expression (Table 1). Then we used the expression profiles of rufipogon or nivara as the outgroup to estimate gene expression changes between Asian domesticated and wild rice. This analysis identified a large portion of the genome, ranging from 1767 to 4489 genes, exhibited differential expression between Asian domesticated and wild rice. Note that several well-characterized genes involved in rice domestication, e.g. *Rc* [32], *gn1a* [33] and *Waxy* [34], are included in the identified DEGs. Interestingly, as a general genome expression trend, we found more genes that were down regulated in cultivated lineages than up regulated (Fig. 2A). The magnitude of change in expression levels was also more dramatic for down-regulated genes than up-regulated genes (Fig. 2B).

To test whether sequences mismatches biased the observed genome expression trend, we selected cultivated rice genes with perfect matches to probes on the microarray and wild rice genes with variable extents of mismatch and performed the identical analysis as shown in Figs. 2A and B. From this analysis, we found that a vast majority of DEGs have a perfect match to the corresponding microarray probe (Fig. S3). Moreover, in this group of genes with perfect matches to the probes, we also observed that in Asian cultivated lineages more genes were down regulated than up regulated (Fig. 2C). The magnitude of expression level changes was also more dramatic for the down-regulated genes than the up-regulated genes (Fig. 2D). Together these results indicate that sequence mismatch does not bias our conclusion regarding the genome expression changes in cultivated rice lineages.

We then compare the overlap of genes expressed differentially between the three Asian cultivated rice and Asian wild rice. There are 823 and 645 genes expressed differentially between all three cultivated rice and *rufipogon* or *nivara*, respectively (Figs. 3A and B). We consider these genes to represent the common differences between Asian cultivated rice and wild rice. We also observed dramatic decrease in genome expression in Asian cultivars relative to Asian wild rice (Figs. 3C and D). For instance, in the comparison between *javanica* and *nivara*, even though the magnitude of expression level changes of down-regulated genes was not significantly larger than that of up-regulated genes (*p*-value=0.3478), the number of down-regulated genes (437 genes) was twice as large as the number of up-regulated genes (208 genes). A total of 17.25% (142 out of 823) genes and 62.45% (514 out of 823) genes are up-regulated and down-regulated, respectively, in all three Asian cultivars when compared to *rufipogon* (Fig. 3E group I and II). Similarly, 23.72% (153 out of 645) genes and 49.92% (322 out of 645) genes are up-regulated and down-regulated, respectively, in all three Asian cultivars when compared to *nivara* (Fig. 3F group I and II). These results further indicate that repression or down-regulation is the predominant trend of genome expression in cultivated rice lineages.

Genes expressed differentially between Asian cultivated rice and Asian wild rice are enriched in biochemical pathways related to energy metabolism

To infer the biological significance of the identified gene expression differences among different rice lineages, all rice genes included in the microarray were analyzed by using the KEGG database [35] and the KOBAS software [36] to identify the particular biochemical pathways which they are involved with. We found that genes expressed differentially between Asian cultivated rice and Asian wild rice were significantly enriched in the pathways related to energy metabolism, such as carbon fixation, starch and sucrose metabolism, the TCA cycle, oxidative phosphorylation, and sugar biosynthesis and interconversion (Table S2).

It's interesting that all three gene sets with differential expression between Asian cultivated rice and Asian wild rice were significantly enriched in starch and sucrose metabolism, carbon fixation, alanine and aspartate metabolism, pyruvate metabolism, and citrate cycle

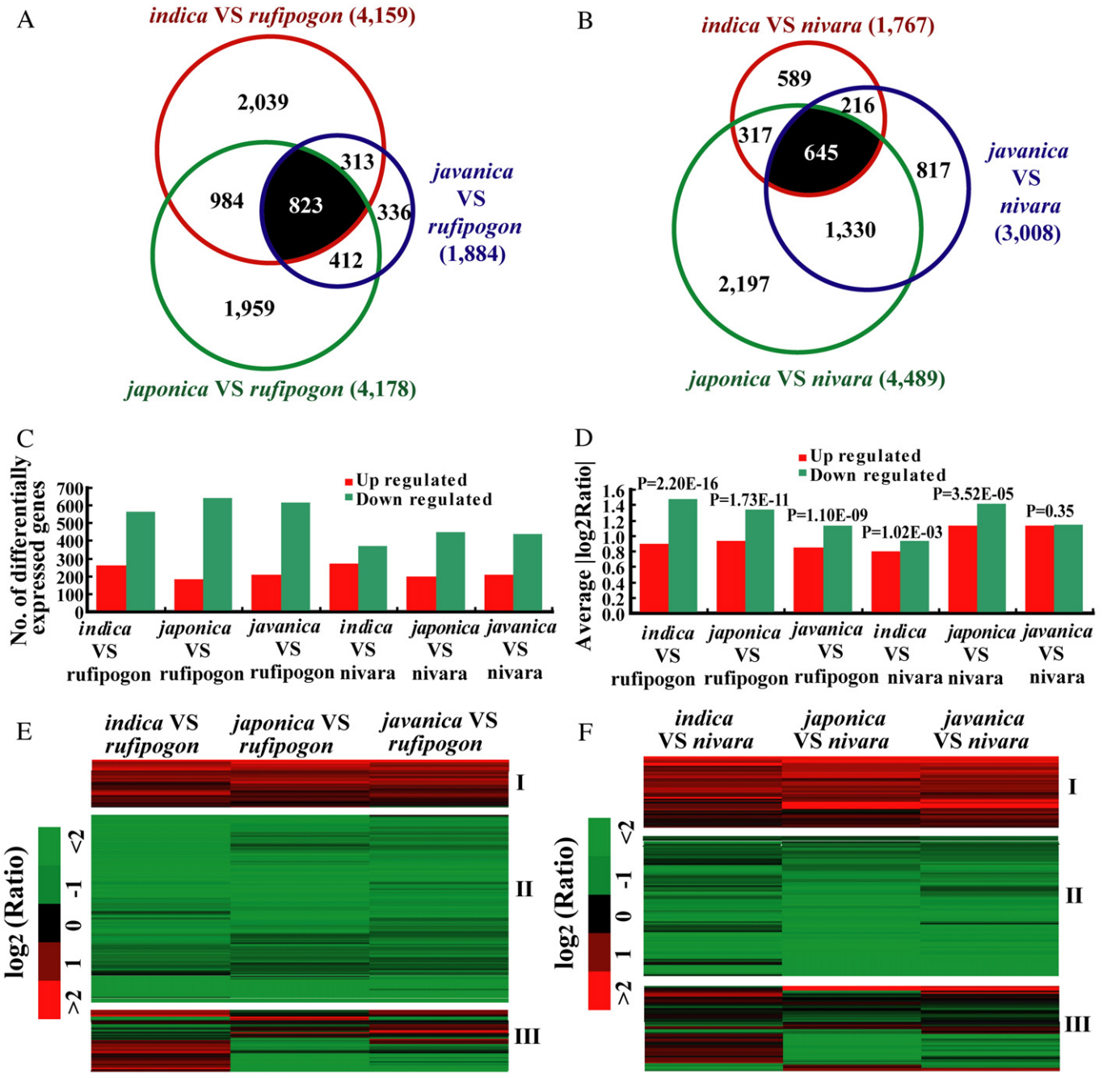

Fig. 3. Analysis of DEGs exhibiting the same expression trend in Asian cultivated rice. A and B) Overlap of genes expressed differentially between Asian cultivated rice and Asian wild rice. C and D) The number (C) and the average expression level change (D) of DEGs that is marked in black in panel A and panel B. The *p*-value from Wilcox test with the null hypothesis that up-regulated genes have the larger mean absolute log2 ratio than down regulated-genes was shown for each comparison. E and F) Hierarchical clustering results of genes marked in black in panel A and panel B.

(pathways marked with *** in Table S2). After identifying the DEGs and their encoded enzymes involved in these pathways, we found that though different genes were selected during the domestication of the three different cultivated rice strains, the enzymes they involved in were almost the same (Fig. 4A). This was especially the case in starch and sucrose metabolism.

We found that the expression levels for genes encoding enzymes related to hydrolyzation was generally lower in the cultivated rice than in the wild rice. Taking the comparison between *javanica* and *nivara* as an example, as shown in Fig. 4B, the expression levels for genes encoding enzymes involved in sucrose-6P hydrolyzation (3.2.1.26), sucrose hydrolyzation (3.2.1.26), cellulose hydrolyzation (3.2.1.4), trehalose hydrolyzation (3.2.1.28), starch hydrolyzation (3.2.1.1 and 3.2.1.2) and maltose hydrolyzation (2.4.1.25) were all reduced in *javanica*. On the other hand, the expression level of enzymes related to amylose and sucrose biosynthesis (2.7.7.27, 2.4.1.11 and 2.4.1.13) was increased in *javanica*. In the TCA cycle, more genes were down regulated in *javanica* than in *nivara* (Fig. S4). Furthermore, comparing *javanica* with *nivara*, down-regulated genes were enriched in the glycolysis pathway while up-regulated genes were enriched in the gluconeogenesis pathway (Fig. S5). These results indicate that the expression of genes related to catabolism was reduced while the expression of genes related to anabolism was induced in the heading-stage panicle tissue of cultivated rice lineages.

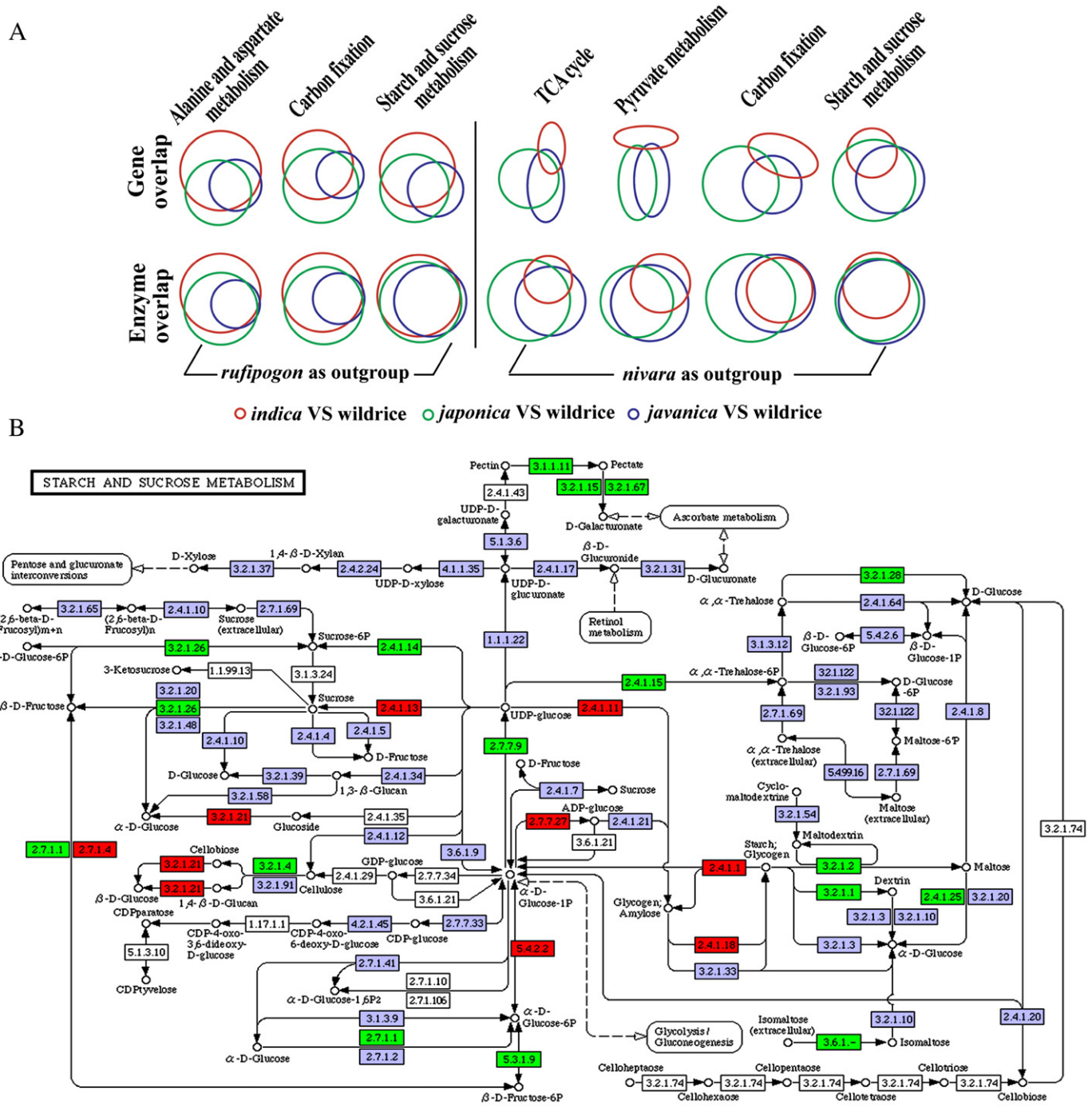

Fig. 4. Analysis of metabolic pathways significantly enriched with DEGs. (A) Overlap of genes and their encoded enzymes expressed differently between Asian cultivated rice and Asian wild rice. (B) The comparison of enzymes encoded by DEGs between *javanica* and *nivara* in the starch and sucrose metabolism pathways. Enzymes highlighted in red are up-regulated in *javanica* than in *nivara*. Enzymes highlighted in green are down-regulated in *javanica* than in *nivara*. Enzymes highlighted in blue represent those that are not differentially expressed between *javanica* and *nivara*.

Candidate genes for phenotypic variation between cultivated and wild rice

Group I and II genes shown in Figs. 3E and F are DEGs that have the same directional change between all three Asian cultivars and the wild rice (Table S3 and Table S4). We postulate that these genes may represent the core gene expression difference related to rice domestication and thus include candidate genes for the phenotypic variation between cultivate and wild rice. For instance, as a trade-off for increase in yield, cultivated cereal grains have up to 50% reduction in the protein content as compared to their wild relatives [37]. Consistent with this observation, we found that genes encoding enzymes in amino acids synthesis, ribosomal proteins, and seed storage proteins are among the group II genes (Figs. 3E and F, Table S3 and S4). To further test this hypothesis, we identified among these genes those that are associated with known functions. Here we show the analysis of genes in the biochemical pathways enriched with DEGs (e.g. genes marked with *** in Table S2) and genes in the transcription factor families enriched with DEGs (Table S5). The resultant candidate genes are shown in Fig. S6. It is worth noting that the well-characterized rice domestication gene Waxy (Os06g04200) [34] is included as one of the candidate genes underlying the phenotypic variations between Asian cultivated rice and the wild rice *rufipogon* (Fig. S6A).

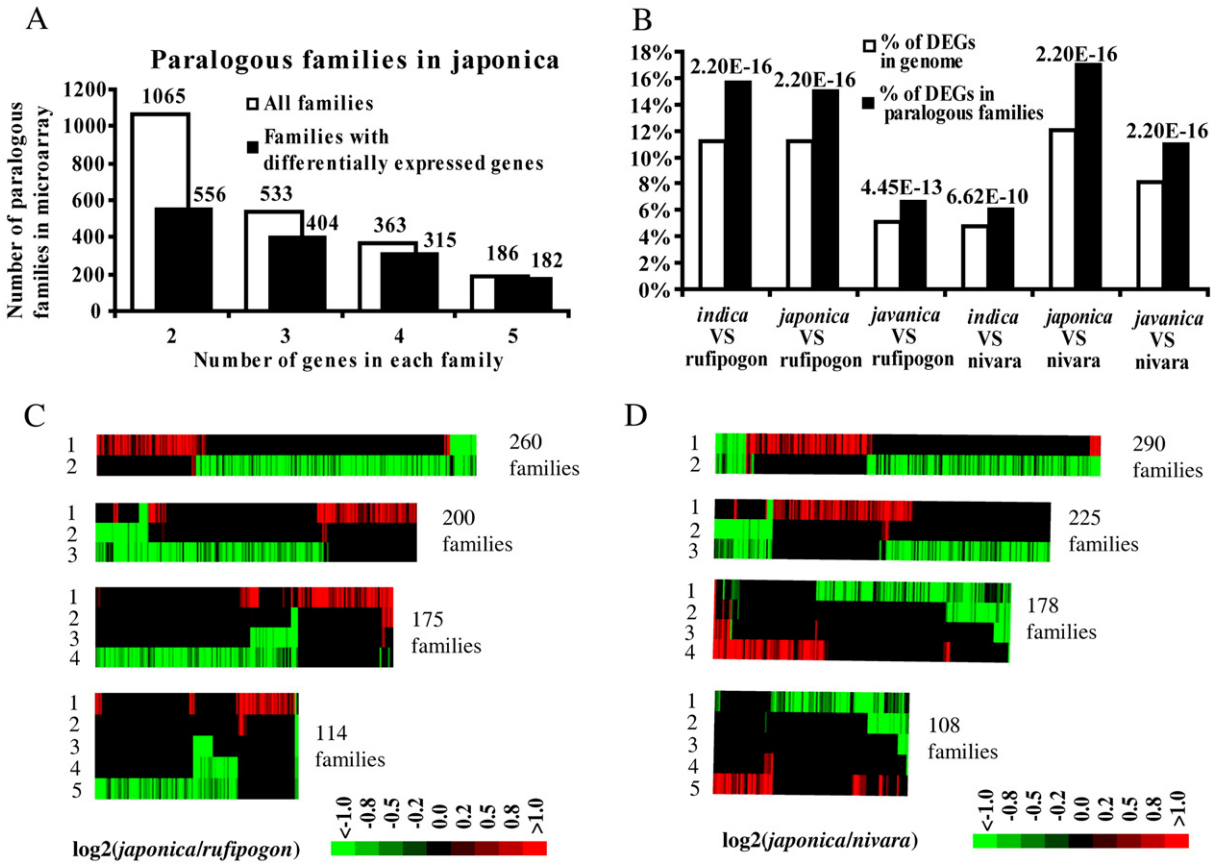

Fig. 5. DEGs in paralogous gene families. (A) Number of *japonica* paralogous families represented in the microarray and families containing DEGs. (B) Genes expressed differentially between different rice lineages show enrichment in paralogous families. The *p* value of the proportion test is shown for each comparison for which the null hypothesis is that the percentage of DEGs in the genome is greater than the percentage of DEGs in paralogous families. C and D) Relative expression level of paralogous families members (only families with 2 to 5 members are shown) that contain DEGs between *japonica* and *rufipogon* (C) or *nivara* (D). Each family member is shown as a bar with each column represents a paralogous family.

Transcription factors regulate the binding of RNA polymerase to target genes. Changes in the expression of transcription factor genes can thus result in expression difference in many genes. Our expression analysis revealed that though transcription factor genes as a whole do not show any enrichment in DEGs relative to the whole genome (Fig. S7), several transcription factor families were significantly enriched in genes expressed differentially between Asian cultivated rice and Asian wild rice (Table S5). The function of many of these transcription factors is related to the regulation of growth or inflorescence and flower development (e.g. ARF, FHA, MADS, ZIM, and C2C2-CO-like). Among these genes, several show the same directional change between all three Asian cultivars and the wild rice (Fig. S6) and are thus worth further investigation aimed at the genetic basis of the phenotypic differences between cultivated and wild rice.

DEGs in paralogous families

In the current TIGR rice genome annotation, a total of 3842 paralogous protein families containing 20,729 proteins were identified in *japonica* based on the existence of protein domains [38]. We found that 13,599 genes from 2670 families were represented by the microarray probes. We found that most individual members, particularly those in families with more than 3 members, have different expression levels as measured by the microarray analysis (Fig. 5A), consistent with the hypothesis that gene family members have adopted new expression patterns after gene duplication. This observation prompted us to investigate the relationship between paralogous families and DEGs among Asian cultivated and wild rice.

The gene families are significantly enriched with DEGs in all comparisons between Asian cultivated and wild rice (Fig. 5B). We then focused on the comparisons between *japonica* and wild rice (e.g. *rufipogon* and *nivara*). There are 4178 and 4489 DEGs between *japonica* and *rufipogon* or *nivara*, respectively (Table 1). Of these, 2049 (*japonica* vs *rufipogon*) and 2320 (*japonica* vs *nivara*) are members of paralogous gene families (Fig. 5B). Thus, approximately half of the DEGs are in paralogous families. Interestingly, we observed that in the vast majority of paralogous families with DEGs between *japonica* and the wild rice, only a single member showed differential expression (Figs. 5C and D). Furthermore, down regulation of expression is more prominent than up regulation, consistent with the general genome expression trend between cultivate and wild rice. Together these results indicate that differential regulation of a specific member of the paralogous gene families is a major driving force for the observed genome expression trend in the examined rice lineages.

Distinct evolutionary forces acted on gene expression and coding sequences

A number of studies have been carried out to test whether gene expression variation in closely related species is under natural selection. In these studies, the linear relationship between species divergence time and the sum of multiple genes' expression differences between species is tested. Those studies led to both positive [31] and negative [10,14] opinions. Since the divergence time among rice species has not been resolved [39], we adapted an alternative method to test this hypothesis. We selected orthologous genes between *indica*

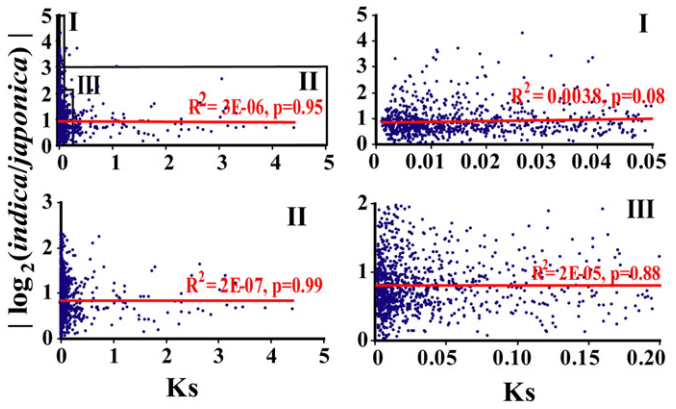

Fig. 6. Correlation of gene expression variation and Ks. *X*-axis denotes the synonymous rate Ks. *Y*-axis denotes the absolute value of log2 ratios of *indica* expression levels and *japonica* expression levels. I, II, III were areas zoomed in from corresponding local area of the upper left figure. In each panel, a variable linear regression was performed and the regressive lines are shown in red.

and *japonica* as our test dataset since their sequence information is complete. We used the synonymous substitution rate Ks and the log2 ratios of expression levels between *indica* and *japonica* orthologous genes as the indicator for gene divergence time and gene expression variation. After analyzing the correlation between gene expression variation and gene divergence time on a gene-by-gene basis, we found no correlation (Fig. 6). This result indicates that gene expression variation in the two tested rice lineages was not linear with their divergence time. Therefore, our finding implied that gene expression was under natural selection [31], but not under neutral evolution [10].

We next examined whether the selection pressure acting on gene expression was different from the selection pressure acting on coding sequences. We performed a test focusing on the DEGs between *indica* and *japonica*. After calculating the $\log_2$ ratios of gene expression level and the ratios of nonsynonymous substitution rates to synonymous rates (Ka/Ks) of DEGs, we found that their correlation was virtually zero (Fig. 7). Chi-square test also showed that Ka/Ks had no effect on expression variation (Ka/Ks <1: p=0.7948; Ka/Ks >1: p=0.8850). Ka/Ks is generally interpreted as the indicator of selective constraint acting on coding sequences. If the expression difference can be considered as an indicator of selective constraint acting on gene expression, our observation suggests that the effects of natural selection on gene coding sequence and gene expression are not statistically correlated. Therefore, we conclude that the selection pressure acting on gene expression is different from that acting on coding sequences in the examined rice lineages.

Discussion

Crop domestication is a long process of selecting traits amenable for cultivation and consumption that resulted in major alterations in plant structure and reproductive physiology [22]. Panicle morphology is one of the main determinants of rice yield [23] and also shows great variations between cultivated rice and wild rice [24]. Compared to wild rice, cultivated rice displays an increased synchronization of panicle formation, more secondary panicle branches [22], and more densely packed panicles that can carry larger numbers of seeds than the wild ancestors [25]. Thus, understanding the genome expression changes in the panicle between cultivated and wild rice will provide insight into the genetic basis of traits associated with the rice domestication syndrome. Not surprisingly, the handfuls of rice genes or QTLs that have been linked to domestication are mainly related to panicle and grain morphology [32–34,40–51].

Domestication traits are often complex phenotypes under the control of numerous genetic components in diverse biochemical and regulatory pathways [22]. Further, many phenotypic changes between closely related species are due to differences in gene expression regulation rather than in the coding sequences [1]. Therefore comparison of the genome expression differences in the rice heading-stage panicle among cultivated and wild rice lineages should provide a stepping stone in uncovering the genetic basis of key phenotypic variations associated with domestication.

We report here a comparative microarray analysis of genome expression of the panicle tissue in the heading-stage in six different rice lineages (Fig. 1). We found that a large proportion of rice genes may have evolved under strong selective constraint which makes their expression levels stable in all six rice lineages. The GO functional enrichment analysis implies that changes in regulation of these genes may be deleterious to plant survival for it will significantly affect plant development, signal regulation, transcription activity and various kinds of plant defense responses (Table S1).

Among the DEGs (Table 1), we found that more genes were down regulated in all three Asian cultivars relative to Asian wild rice. Furthermore, the expression levels changed more in down-regulated genes (Fig. 2). To study the gene expression differences at a systematic level, we then identified all metabolic pathways in which the DEGs involved. As shown in Fig. 4A, though different genes were selected during the domestication of the three different cultivated rice strains, the enzymes these genes encode were almost the same. This was especially true in starch and sucrose metabolism, probably because starch is one of the key components of cereal grains and thus a target for selection during domestication [52]. Metabolic pathway analysis also showed that DEGs between Asian cultivated rice and Asian wild rice were significantly enriched in the pathways related to energy metabolism (Table S2). Further analysis indicates that the expression of genes related to catabolism was reduced while the expression of genes related to anabolism was induced in the heading-stage panicle of cultivated rice lineages (Fig. 4). This trend in gene expression changes is clearly associated with the selection for higher yields in rice domestication.

So far the well-characterized genes related to crop domestication are predominantly transcription factors [37], regulating phenotypic variation ranging from inflorescence structure to seed shattering. However, transcription factors as a whole did not show any enrichment in DEGs between cultivated and wild rice (Fig. S7). The GO term analysis of genes that expressed constantly among all six rice lineages also showed that expression levels of genes related to transcription activity were relatively stable (Table S1). These observations are consistent with findings in *Drosophila*, in which the expression levels of transcription factors were found to evolve slower than other types of genes [14,15].

Nonetheless, we found that several transcription factor families were significantly enriched in genes expressed differentially between Asian cultivated rice and wild rice. Three families of transcription factors were related to the regulation of inflorescence and flower development. They belong to a MADS family involved in floral organ identity determination and flower formation [53,54], a ZIM family

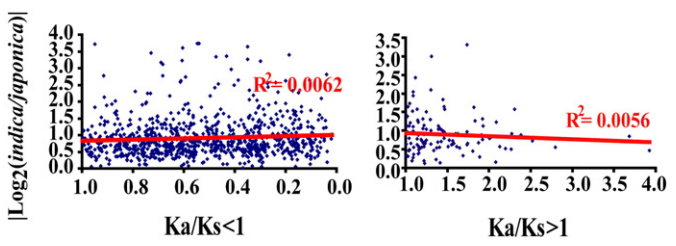

Fig. 7. Correlation of expression variation and Ka/Ks. *X*-axis denotes the Ka/Ks of genes expressed differentially between indica and *japonica*. *Y*-axis denotes the absolute log2 ratios of *indica* and *japonica* expression level. A variable linear regression was performed and the regressive lines are shown in red.

involved in inflorescence and flower development [55], and a C2C2-CO-like family involved in flowering control [56,57]. This finding is consistent with the fact that cultivated rice displays increased synchronization of panicle formation and panicle development than its wild ancestors. Two families, ARF and FHA, were related to the regulation of cell cycle and growth [58,59], consistent with the fact that cultivated rice displays larger panicle size than its wild ancestors. Our results also suggest that the phenotypic changes of heading-stage panicles between cultivated and wild rice may be mainly affected by a relatively small number of transcription factors. Focusing on these transcription factors may help to identify new candidates underlying the important phenotypic variations of heading-stage panicles between cultivated and wild rice.

In addition to revealing the genome expression differences between cultivated rice and wild rice, our study also explored the general trends of rice gene expression evolution. First, we found that approximately three quarters of rice genes have constant expression level across all six examined rice lineages. Previous gene expression evolution study from other model organisms also showed that stabilizing selection, which was a type of natural selection that decreased genetic diversity and made population stabilizing on a particular trait, was likely to be a dominant mode of gene expression evolution [11,14,31]. Though we could not test definitely whether these rice genes were under stabilizing selection from our data, the fact that expression levels of many genes have kept more or less constant is consistent with the previous study [11,14,31].

Second, we found no correlation between gene expression variation and the synonymous substitution rate Ks for genes expressed differentially between *indica* and *japonica* (Fig. 6). As Ks is usually used as the indicator for gene divergence time, this result indicates that gene expression variation in different rice lineages was not linear with their divergence time. Furthermore, we considered the expression difference and the Ka/Ks ratio as indicator of selective constraint acting on gene expression and the coding sequences, respectively. Our observation suggests that the effects of natural selection on gene coding sequence and gene expression were different, since the correlation was very weak (Fig. 7). This result is in agreement with a previous analysis of the syntenic relationship between barrel medic and soybean in the genome regions surrounding two soybean cyst nematode resistance loci in which the identified syntenic regions showed significant differential gene expression [60]. Thus our result is consistent with the hypothesis put forth by Mary-Claire King and Allan Wilson three decades ago that genes evolve individually at the fronts of coding sequences and regulation of expression [1].

Finally, we found that approximately half the DEGs are in the paralogous gene families (Figs. 5C and D). In the vast majority of the families with DEGs, only a single member showed differential expression (Figs. 5C and D). Furthermore, down regulation of expression is more prominent than up regulation in these families, consistent with the general genome expression trend between cultivate and wild rice. These observations indicate that differential regulation of a specific member of the paralogous gene families is a major driving force for the observed genome expression trend in the examined rice lineages. These results are consistent with phylogenetic analyses showing that the evolutionary rate of repression or loss of gene expression regions significantly outpaces the rate of activation or gain [61]. Furthermore, these results can be viewed as support to the proposal that loss of gene function may represent a common evolutionary response to the short-term selections [62].

In summary, we employed microarray experiments as a new approach to generate a global view of genome expression trends in the heading-stage panicle in six cultivated and wild rice lineages. We observed that repression of genome expression (i.e. reduction of the number of expressed genes and the expression level of expressed genes) represents the dominant trend in the heading-stage panicle as a result of rice domestication. This genome expression trend appears to be accomplished in part by the differential expression of a single member from a large number of paralogous families. Examination of the pathways enriched with DEGs between cultivated and wild rice showed that the expression level of genes related to catabolism was repressed and that of genes related to anabolism increased in the cultivars. We also found that several transcription factor families related to the regulation of flower development and cell growth were significantly enriched in DEGs. Further studies targeting these transcription factors may help to identify candidate genes controlling the desired domestication traits. Finally, we found that distinct evolutionary forces affected gene expression and coding sequences differentially. These findings demonstrated the importance of studying genetic differences for phenotypic variations at the genome level and provide a stepping stone in uncovering the genetic basis of phenotypic variations between cultivated and wild rice.

Materials and methods

Plant materials

Three subspecies of Asian cultivated rice, *Oryza sativa* ssp. *indica* (9311, China), *O. sativa* ssp. *japonica* (Nipponbare, Japan), and *O. sativa* ssp. *javanica* (Sipak, Indonesia) and one accession of the African cultivated rice, *O. glaberrima* (Macina, Mali) were used in this study. We also sampled two wild species that are closely related to the Asian cultivated rice, i.e., *O. rufipogon* (IRRI Accession no. 106161, Laos) and *O. nivara* (IRRI Accession no. 106061, India). Panicles grown in a controlled growth chamber at 28 °C were harvested at heading stage, frozen in liquid nitrogen immediately and then kept at −80 °C before RNA extraction.

Microarray analysis

The previously described rice 70-mer oligonucleotide set [27] that contains 36,926 unique genes was used in this study. Probe preparation, labeling, and microarray hybridization were carried out as previously described [27]. After manual removal of spots with aberrant morphology, microarray spot intensity signals were acquired using the Axon GenePix Pro 5.0 software without correction for background. The raw signal intensities were normalized by Lowess normalization followed with quantile normalization for the further analysis. The raw and normalized microarray data are available in the NCBI Gene Expression Omnibus under the serials GSE11712.

To identify genes whose expression levels remained constant in all six rice lineages, we used the ANOVA *F*-test with the null hypothesis that the mean expression levels of each rice lineages were equal. Genes with the same expression in all lineages were determined by the ANOVA *F*-test and a Bonferroni multiple adjustment with p-value>0.05 using the MAANOVA package for *R*. To reduce false positive genes, we selected genes with p-value>0.7 for the stringent analysis. Further GO functional enrichment analysis was performed on the dataset with p-value>0.7.

To identify DEGs between any two rice lineages, we used the ANOVA *T*-test with the null hypothesis of no expression difference between the two lineages. Differential expression was determined by the ANOVA *T*-test and a Bonferroni multiple adjustment with $p<0.05$ using the MAANOVA package for *R*. Cluster analysis was applied to genes showing differential expression between three Asian cultivated rice and wild rice. Hierarchical clustering with the complete linkage was performed using the software Cluster [63] and visualized by the Java Treeview program.

RT-PCR analysis

Total RNA from heading-state panicles used in microarray analysis were treated with RNase-free DNase (Promega) and cDNA was

synthesized with the SuperScript II First-Strand cDNA Synthesis kit (Invitrogen). Genes met with the following criteria were selected for RT-PCR experiment. First, the genes should be contained in BAC clone sequences from *rufipogon*, *nivara* and *glaberrima*. Second, the overlap of *indica* gene model, *japonica* gene model, *rufipogon* BAC clone sequences, *nivara* BAC clone sequences and *glaberrima* BAC clone sequences should be longer than 150nt with ≥90% sequence identity. Then the primers were determined based on the common sequences. PCR was performed using general standard techniques and an *O. sativa* actin gene was used as control.

Pathway and GO category analysis

We used the software KOBAS (KEGG Orthology Based Annotation System [36]) to identify biochemical pathways related to DEGs between Asian cultivated rice and Asian wild rice and to calculate the statistical significance of each pathway. We used KOBAS to assign rice genes to pathways by matching them to similar genes (as determined by BLAST similarity search with cutoff *e*-values $<1e^{-5}$, rank <10, and sequence identity >30%) in known *Arabidopsis* pathways in the KEGG database. We also manually reviewed all identified pathways for quality control.

Next, we used KOBAS to rank pathways by the *p* value, which is designed to test whether data from a particular pathway fits the null hypothesis or the alternative hypothesis defined as

$$H_0 : p_0 \le p_1$$
$$H_1 : p_0 > p_1$$

where $p_0 = m/M$, $p_1 = n/N$, m is the number of DEGs mapped to the pathway under investigation, M is the number of all DEGs with KOBAS annotation, n is the number of all genes mapped to the selected pathway, and N is the total number of genes with KOBAS annotation. The p value of a particular pathway corresponds to a test statistic following a hypergeometric distribution:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{M}{i}\binom{N-M}{n-1}}{\binom{N}{n}}$$

Pathways with P values<0.05 were considered statistically significant.

We first downloaded the rice GO annotation (version 4) from TIGR database [38]. Then the same statistic model described at pathway analysis was used to identify enriched GO terms. Here, m is the number of genes with constant expression levels that are annotated with the given GO term, M is the number of all genes with constant expression levels with GO annotation, n is the number of all genes annotated with the given GO term, and N is the total number of genes with GO annotation. GO terms with adjusted P values<0.05 by Bonferroni's correction for multiple tests were considered statistically significant.

Gene family and transcription factor genes

We downloaded the *japonica* paralogous families from release 4 of TIGR rice database [38]. We identified gene families with at least one member that belongs to DEGs. For the analysis of transcription factor genes, we first retrieved the list of rice transcription factors and their corresponding families from the DRTF database using the *japonica* data [64]. Then we calculated the proportion of transcription factors in DEGs. The proportions of transcription factors in whole genome were all calculated from *japonica* dataset. The *p* values of the hypergeometric tests, which the null hypothesis was that the proportion of transcription factors in DEGs was not smaller than that in whole genome, were also calculated. Finally, we identified transcription factor families that were enriched in DEGs by using the same statistical model described for pathway analysis.

Ka/Ks and chi square analysis

We first selected orthologous gene pairs between *indica* and *japonica* as previously described [27]. The Ka/Ks of each orthologous pair was calculated with the NG method by using the PAML yn00 package [65]. To test whether Ka/Ks has an effect on gene expression difference, we performed scatter plot of Ka/Ks and log2 *indica*/*japonica* expression ratio. We divided the entire scatter plot to *n* by *n* zones with equal area for each zone and counted the number of genes fall in each zone to generate a contingency table with *n* columns by *n* rows. Then, Pearson's chi-square test with the null hypothesis that Ka/Ks has no effect on expression difference was carried out with 1,000,000 times Monte Carlo simulation by using the chisq.test() function in the *R* statistic environment.

Acknowledgments

We thank Louis Tao (center for bioinformatics in Peking University) for critical reading of the manuscript, and Manyuan Long (Chicago University) for discussion of evolutional concept. The study was supported by “the 863 rice functional genomics program” from the Ministry of Science and Technology of China and the National Institute of Biological Sciences, Beijing.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2008.10.005.

References

[1] M.C. King, A.C. Wilson, Evolution at two levels in human and chimpanzees, Science 188 (1975) 107–116.
[2] M.V. Rockman, et al., Ancient and recent positive selection transformed opioid cis-regulation in humans, PLoS Biol. 3 (2005) e387.
[3] P.D. Keightley, M.J. Lercher, A. Eyre-Walker, Evidence for widespread degradation of gene control regions in hominid genomes, PLoS Biol. 3 (2005) e42.
[4] A.P. Gasch, et al., Conservation and evolution of cis-regulatory systems in ascomycete fungi, PLoS Biol. 2 (2004) e398.
[5] A.M. Moses, D.Y. Chiang, D.A. Pollard, V.N. Iyer, M.B. Eisen, MONKEY: identifying conserved transcription-factor binding sites in multiple alignments using a binding site-specific evolutionary model, Genome Biol. 5 (2004) R98.
[6] W.J. Dickinson, On the architecture of regulatory systems: evolutionary insights and implications, Bioessays 8 (1988) 204–208.
[7] M.Z. Ludwig, et al., Functional evolution of a cis-regulatory module, PLoS. Biol. 3 (2005) e93.
[8] M.Z. Ludwig, C. Bergman, N.H. Patel, M. Kreitman, Evidence for stabilizing selection in a eukaryotic enhancer element, Nature 403 (2000) 564–567.
[9] Y. Gilad, A. Oshlack, S.A. Rifkin, Natural selection on gene expression, Trends Genet. 22 (2006) 456–461.
[10] P. Khaitovich, et al., A neutral model of transcriptome evolution, PLoS Biol. 2 (2004) E132.
[11] P. Khaitovich, et al., Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees, Science 309 (2005) 1850–1854.
[12] W. Enard, et al., Intra- and interspecific variation in primate gene expression patterns, Science 296 (2002) 340–343.
[13] M. Caceres, et al., Elevated gene expression levels distinguish human from non-human primate brains, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 13030–13035.
[14] S.A. Rifkin, J. Kim, K.P. White, Evolution of gene expression in the *Drosophila* melanogaster subgroup, Nat. Genet. 33 (2003) 138–144.
[15] S.A. Rifkin, D. Houle, J. Kim, K.P. White, A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression, Nature 438 (2005) 220–223.
[16] B. Lemos, C.D. Meiklejohn, M. Caceres, D.L. Hartl, Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories, Evolution Int. J. Org. Evolution 59 (2005) 126–137.
[17] D.R. Denver, et al., The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*, Nat. Genet. 37 (2005) 544–548.
[18] J. Yu, et al., The genomes of *Oryza sativa*: a history of duplications, PLoS Biol. 3 (2005) e38.
[19] J. Yu, et al., A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*), Science 296 (2002) 79–92.
[20] S.A. Goff, et al., A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*), Science 296 (2002) 92–100.
[21] J.S. Ammiraju, et al., The *Oryza* bacterial artificial chromosome library resource: construction and analysis of 12 deep-coverage large-insert BAC libraries that represent the 10 genome types of the genus *Oryza*, Genome Res. 16 (2006) 140–147.

[22] M.J. Kovach, M.T. Sweeney, S.R. McCouch, New insights into the history of rice domestication, Trends Genet. 23 (2007) 578–587.
[23] I. Furutani, S. Sukegawa, J. Kyozuka, Genome-wide analysis of spatial and temporal gene expression in rice panicle development, Plant J. 46 (2006) 503–511.
[24] T. Sang, S. Ge, Genetics and phylogenetics of rice domestication, Curr. Opin. Genet. Dev. 17 (2007) 533–538.
[25] M. Sweeney, S. McCouch, The complex history of the domestication of rice, Ann. Bot. (Lond) 100 (2007) 951–957.
[26] Y. Wang, J. Li, The plant architecture of rice (*Oryza sativa*), Plant Mol. Biol. 59 (2005) 75–84.
[27] L. Ma, et al., A microarray analysis of the rice transcriptome and its comparison to *Arabidopsis*, Genome Res. 15 (2005) 1274–1283.
[28] C. Cheng, et al., Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs, Mol. Biol. Evol. 20 (2003) 67–75.
[29] H. Morishima, Y. Sano, H.I. Oka, Evolutionary Studies in Cultivated Rice and Its Wild Relatives, Oxford Surveys in Evolutionary Biology, 1992.
[30] H.I. Oka, Origin of Cultivated Rice, Scientific Societies Press/Academic Press, Tokyo, Japan, 1988.
[31] Y. Gilad, A. Oshlack, G.K. Smyth, T.P. Speed, K.P. White, Expression profiling in primates reveals a rapid evolution of human transcription factors, Nature 440 (2006) 242–245.
[32] M.T. Sweeney, M.J. Thomson, B.E. Pfeil, S. McCouch, Caught red-handed: Rc encodes a basic helix–loop-helix protein conditioning red pericarp in rice, Plant Cell 18 (2006) 283–294.
[33] M. Ashikari, et al., Cytokinin oxidase regulates rice grain production, Science 309 (2005) 741–745.
[34] K.M. Olsen, et al., Selection under domestication: evidence for a sweep in the rice waxy genomic region, Genetics 173 (2006) 975–983.
[35] M. Kanehisa, et al., From genomics to chemical genomics: new developments in KEGG, Nucleic Acids Res. 34 (2006) D354–D357.
[36] X. Mao, T. Cai, J.G. Olyarchuk, L. Wei, Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary, Bioinformatics 21 (2005) 3787–3793.
[37] J.F. Doebley, B.S. Gaut, B.D. Smith, The molecular genetics of crop domestication, Cell 127 (2006) 1309–1321.
[38] S. Ouyang, et al., The TIGR Rice Genome Annotation Resource: improvements and new features, Nucleic Acids Res. 35 (2007) D883–D887.
[39] Q. Zhu, S. Ge, Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes, New Phytol. 167 (2005) 249–265.
[40] C. Li, A. Zhou, T. Sang, Rice domestication by reducing shattering, Science 311 (2006) 1936–1939.
[41] Z. Lin, et al., Origin of seed shattering in rice (*Oryza sativa* L.), Planta 226 (2007) 11–20.
[42] S. Konishi, et al., An SNP caused loss of seed shattering during rice domestication, Science 312 (2006) 1392–1396.
[43] S. Yamanaka, I. Nakamura, K.N. Watanabe, Y. Sato, Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice, Theor. Appl. Genet. 108 (2004) 1200–1204.
[44] C. Fan, et al., GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein, Theor. Appl. Genet. 112 (2006) 1164–1171.
[45] L.M. Bradbury, T.L. Fitzgerald, R.J. Henry, Q. Jin, D.L. Waters, The gene for fragrance in rice, Plant Biotechnol. J. 3 (2005) 363–370.
[46] X.J. Song, W. Huang, M. Shi, M.Z. Zhu, H.X. Lin, A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase, Nat. Genet. 39 (2007) 623–630.
[47] H.W. Mei, et al., QTLs influencing panicle size detected in two reciprocal introgressive line (IL) populations in rice (*Oryza sativa* L.), Theor. Appl. Genet. 112 (2006) 648–656.
[48] J. Yamagishi, N. Miyamoto, S. Hirotsu, R.C. Laza, K. Nemoto, QTLs for branching, floret formation, and pre-flowering floret abortion of rice panicle in a temperate japonica×tropical japonica cross, Theor. Appl. Genet. 109 (2004) 1555–1561.
[49] C. Li, A. Zhou, T. Sang, Genetic analysis of rice domestication syndrome with the wild annual species, *Oryza nivara*, New Phytol. 170 (2006) 185–193.
[50] L. Xiong, K. Liu, K. Dai, C. Xu, Q. Zhang, Identification of genetic factors controlling domestication-related traits for rice using an F2 population of a cross between *Oryza sativa* and *O. rufipogon*, Theor. Appl. Genet. 98 (1999) 243–251.
[51] W. Cai, H. Morishima, QTL clusters reflect character associations in wild and cultivated rice, Theor. Appl. Genet. 104 (2002) 1217–1228.
[52] S.R. Whitt, L.M. Wilson, M.I. Tenaillon, B.S. Gaut, E.S.T. Buckler, Genetic diversity and selection in the maize starch pathway, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 12959–12962.
[53] E.S. Coen, E.M. Meyerowitz, The war of the whorls: genetic interactions controlling flower development, Nature 353 (1991) 31–37.
[54] J. Nam, C.W. dePamphilis, H. Ma, M. Nei, Antiquity and evolution of the MADS-box gene family controlling flower development in plants, Mol. Biol. Evol. 20 (2003) 1435–1447.
[55] A. Nishii, et al., Characterization of a novel gene encoding a putative single zinc-finger protein, ZIM, expressed during the reproductive phase in *Arabidopsis thaliana*, Biosci. Biotechnol. Biochem. 64 (2000) 1402–1409.
[56] S. Griffiths, R.P. Dunford, G. Coupland, D.A. Laurie, The evolution of CONSTANS-like gene families in barley, rice, and *Arabidopsis*, Plant Physiol. 131 (2003) 1855–1867.
[57] S. Laubinger, et al., *Arabidopsis* SPA proteins regulate photoperiodic flowering and interact with the floral inducer CONSTANS to regulate its stability, Development 133 (2006) 3213–3222.
[58] G. Zhu, et al., Two yeast forkhead genes regulate the cell cycle and pseudohyphal growth, Nature 406 (2000) 90–94.
[59] T.J. Guilfoyle, G. Hagen, Auxin response factors, Curr. Opin. Plant Biol. 10 (2007) 453–460.
[60] L. Li, et al., Transcriptional analysis of highly syntenic regions between *Medicago truncatula* and *Glycine max* using tiling microarrays, Genome Biol. 9 (2008) R57.
[61] T.H. Oakley, B. Ostman, A.C. Wilson, Repression and loss of gene expression outpaces activation and gain in recently duplicated fly genes, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 11637–11641.
[62] M.V. Olson, When less is more: gene loss as an engine of evolutionary change, Am. J. Hum. Genet. 64 (1999) 18–23.
[63] M.B. Eisen, P.T. Spellman, P.O. Brown, D. Botstein, Cluster analysis and display of genome-wide expression patterns, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 14863–14868.
[64] G. Gao, et al., DRTF: a database of rice transcription factors, Bioinformatics 22 (2006) 1286–1287.
[65] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (1986) 418–426.